This pre-print was published as

Moreno, A.I. (2008). The importance of comparable corpora in cross-cultural studies. In Ulla Connor, Ed Nagelhout, and William Rozycki (eds), *Contrastive Rhetoric: Reaching to Intercultural Rhetoric*, pp. 25-41. Amsterdam: John Benjamins.

The importance of comparable corpora in cross-cultural studies

Ana I. Moreno

University of León

Abstract

This chapter shows the importance of comparing corpora that are really comparable. The chapter conceives of texts as exemplars of situated genres and acknowledges that the rhetorical and discourse configuration of texts vary as a function of the contextual factors in which texts are situated. It argues that corpora may be considered equivalent (or similar to the maximum degree) across cultures to the extent that the text exemplars are similar in all of the relevant contextual factors. It concludes that cross-cultural corpora designs should attempt to control statistically as many of the relevant contextual factors as possible. If not, it may not be possible to say anything reliable about the possible effect of the language/culture factor on texts. Instead, possible differences found may be due to uncontrolled contextual variables.

Keywords

Cross-cultural, comparable corpora, contextual factors, confounding variables, culture, corpus linguistics, discourse analysis, contrastive rhetoric

The importance of comparable data in cross-cultural studies

Ana I. Moreno

University of León

Introduction

It has been four decades since Kaplan (1966) proposed the idea that the rhetorical structures of texts in different languages might vary greatly, and that such variation should be taken into account in language teaching programs. He also suggested that these differences in writing across cultures may reflect different writing conventions and, in an attempt to revise his initial notion – which was severely criticized –Kaplan (1987) later on suggested that these differences in writing do not necessarily reflect different patterns of thinking that are acquired, but are more likely to reflect cultural and educational training factors which help to shape the writing conventions that are learned in a culture.

Given the tremendous research activity driven by his basic insight that "writing is culturally influenced in interesting, and complex, ways" (Connor, 2002, p. 495), contrastive rhetoric (CR) scholars should be convinced that this field of enquiry does not need further justification. What they need to do now is whatever is necessary to make their research more and more rigorous, reliable, and explanatory. One thing that can be done in this direction is to make sure they are comparing what is comparable across cultures. The purpose of this chapter is to reflect on

how important this methodological requirement is.

It is first convenient to distinguish between different types of questions that have been considered by CR researchers. A clear distinction is considered crucial because the type of question will ultimately dictate the sources and type of data that should be compared, and which therefore need to be comparable.

Contrastive Rhetoric Questions

1. Whether the imputed cross-cultural differences in the rhetorical configuration of texts actually exist,
2. If they exist, which cultural or educational factors may help to account for such differences (e.g., values, norms, learning processes and educational trends),
3. Which precise difficulties with discourse structure and other rhetorical features do second language learners from a given non-English writing culture experience when writing in English as an L2,
4. Whether difficulties experienced with discourse structure and other rhetorical features by L2 learners of English are attributable to interference (or negative transfer) from the first language.

Table 1 sketches the types of data that should be used to answer each of these questions:

Table 1. Types of comparable data used to answer CR questions

Question type	Types of comparable data
Question 1	Original texts written in English as L1 and ***equivalent*** original texts written independently in the other language as L1
Question 2	Documentation/Information provided by writers/readers/other participants involved in the process of writing, interpreting and learning to write and interpret original texts written in English as L1 and similar information provided by similar participants in similar processes in relation to ***equivalent*** original texts written independently in the other language as L1
Question 3	Original texts written in English as L2 and ***equivalent*** original texts written independently in English as L1. Information about the process of writing and learning to write these texts both as L1 and as L2
Question 4	The results obtained in answer to question 3 (the difficulties experienced) and ***relevant*** results obtained in answer to questions 1 (the differences observed) and 2 (the reasons for these differences)

To answer question 1, CR studies will need to identify areas of difference (and similarity) in equivalent rhetorical/discourse features between original texts written in English as L1 and equivalent original texts written independently in the other language as L1. Accordingly, translations or English as L2 texts would not be valid to answer this question. As Reid (1988) argues, texts written by nonnative speakers – whether L2 texts or translations – do not constitute "a sufficient data sample for valid analysis because they use second language texts to investigate first language rhetorical patterns" (Reid, 1988, p. 19).

To answer question 2, CR studies need to go beyond texts and look contrastively at the two writing cultures, big and small, that have contributed to the L1 texts. This would involve obtaining information from documents and from the writers, readers, and other participants involved in the process of writing, interpreting, and learning to write and interpret the type of texts under research, by means of questionnaires, interviews, and so on.

To answer question 3, CR studies need to identify non-English rhetorical features in original texts written in English as L2 by referring to equivalent discourse and rhetorical features used in equivalent original texts written independently in English as L1 (not only with reference to judgments by native speakers' of English such as those by teachers of English as L2, but with reference to what actually happens in equivalent non-expert texts written in English as L1). They also need information about the process of writing and learning to write these texts in each writing culture independently.

To answer the final question (question 4) CR studies need to compare the results obtained in answer to questions 1 and 2 to the relevant results obtained in answer to question 3. This will allow us to check whether, for instance, the differences identified in answer to question 1 could explain the non-English discourse and rhetorical features identified in answer to question 3.

The main distinguishing feature of CR is that it has broadened the scope of its analyses to incorporate text-linguistic phenomena that go beyond the sentence level. Rather than focus on the lexico-grammatical resources employed by different languages to create meaning, it has concentrated on the text features that help writers of different cultures to negotiate meaning, both on the textual and the interpersonal plane. To this aim, CR has been more interested in finding out how writers of different writing cultures organize their texts into coherent meaningful units to accomplish their communicative purposes. So, for instance, studies have looked at paragraph organization and macro-structures of different kinds. They have also looked at how information flows effectively across sentences. CR has also been concerned with how writers show their attitudes towards their own ideas and their readers, how much conviction and commitment to their ideas writers feel it is appropriate to display depending on the communicative situation, and so on.

In Table 2.1, the texts used as sources of data for the comparison should be *equivalent*, or comparable. Although it sounds rather obvious, meeting this methodological requirement is crucial for any study that aims to contribute valid knowledge to this discipline. As Connor and Moreno (2005) have recently argued, this requirement to use comparable data should be met, to start with, at two basic stages of the research: one, in selecting texts to build parallel corpora; and two, in identifying equivalent

textual concepts to be examined in the corpora. Due to limitations of space, I will focus on the first of these two stages: the selection of parallel corpora.

Definition of Parallel Corpora

A corpus is defined as a sample of texts which may be considered representative of the population of texts which it intends to represent. For example, Moreno's (1996, 1997, 1998) corpus comprised two sets of 36 research articles from Business and Economics, one in English, the other in Spanish. Each of these samples was meant to represent the population of research articles on Business and Economics in each language.

Parallel corpora are defined as sets of comparable original texts written independently in two or more languages (Connor & Moreno, 2005, p. 155). The notion of comparability is equated to the concept of *equivalence* and is crucial for designing corpora for CR studies.

The Concept of Equivalence in Parallel Corpus Designs

The concept of equivalence has been widely used in Contrastive

Analysis and Translation Theory. It helps contrastive researchers to establish a valid criterion of comparison between corpora. One important development in this respect has been that the original conception of equivalence as *identity* is giving way to the conception of equivalence as *maximum similarity* (Chesterman, 1998). That is, for two corpora to be considered as equivalent they do not need to be exactly the same but similar to the maximum degree. But we still face the problem that judgments about what constitutes maximum similarity and how it is to be measured are relative, i.e. they depend on the assessors. So definitions of equivalence (or maximum similarity) will be relative to the theoretical framework in which they are made.

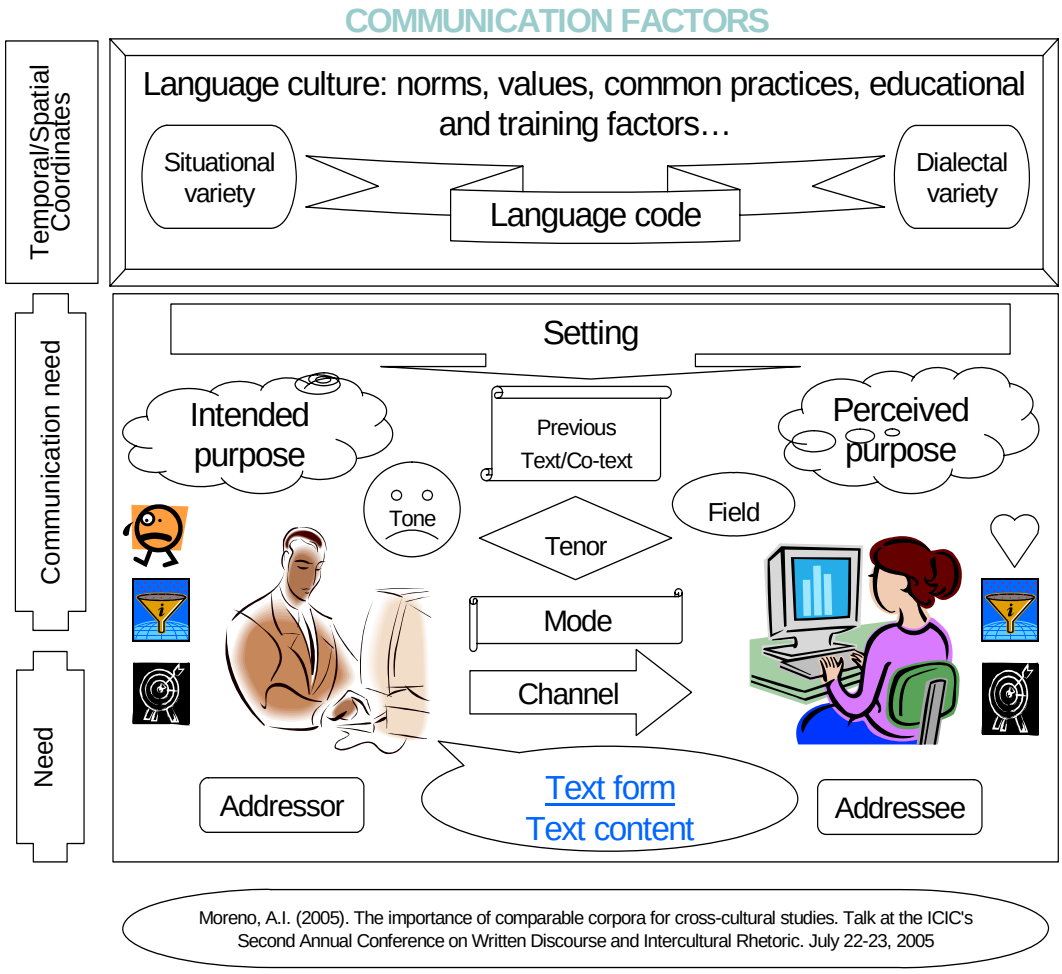
This chapter addresses how this requirement of maximum similarity could be met in cross-cultural studies that draw on one particular theoretical framework, genre theory (Swales, 1990). According to this framework, it is theoretically plausible to conceive of texts as exemplars of situated genres. This means acknowledging that the rhetorical and semantico-linguistic configuration of texts – what is known as *text form* and *content* – vary as a function of the context in which discourse is situated. If we make our concept of maximum similarity relative to this theoretical conception of texts, we could then say that two corpora are equivalent (or similar to the maximum degree) to the extent that the text exemplars contained in them may be considered similar in all

relevant contextual factors. It is therefore necessary to clarify what is meant by *contextual factors* in a general model of communication, applicable to contrastive studies.

A Model of the Communication Process

The concept of *context* of communication is considered from a socio-cultural and cognitive perspective in Figure 1.

Figure 1. Contextual factors affecting cross-cultural communication



Not only does the context comprise relevant information about the immediate physical environment (*setting*) in which communication takes place, the relevant *co-text,* and *other* relevant *texts*, but it also comprises relevant information about the emotional state of the participants (their mood), their goals/needs, their expectations about the future, their anecdotal memories, their beliefs, their general cultural assumptions and previous knowledge and experience of the world and texts, and their mutual knowledge about all these things. According to Relevance Theory, all these factors play a role in the interpretation of texts (Sperber & Wilson, 1986, p. 15). From the point of view of text creation, all of these factors are also important because they may and do have an influence on the form and content of the text.

Figure 1 proposes a model of the communication process which takes most of these factors into account. This model draws on the well-known work by Jakobson (1960), Hymes (1962) – both quoted in Cook (1989) – Halliday and Hasan (1989, p. 12), and Sperber and Wilson (1986). Strictly speaking, it is not really a model; it only includes the elements or components of the system but not the relationships between the different elements, and more elements could still be incorporated. However, it does show the complexity of the communication process. It is hoped that, as research advances in this line, it will be possible to complete the model in terms of the relationships between the different

elements and how each affects the others.

The model presupposes a *communication need* on the part of the writer, which takes place in a given *temporal* and *spatial coordinate* in the history of human relations as a result of some *need* the person wants to satisfy. Needs may be of various kinds: physiological, security, social, esteem, self-actualization – cf. Maslow's (1954) book, *Motivation and Personality* – improving one's representation of the world – cf. Sperber and Wilson (1986) – and so on.

Perhaps the most important element of the communication process is the formulation of a communicative *purpose*, because this is what drives the addressor to enter a given communication process and determines the selection of the different options available to achieve his/her goal. Take, for instance, the communicative purpose of applying for a job in a company. To achieve that purpose, the *addressor* needs to determine the right person to communicate with – that is, the *addressee* (status within the company, sex, age, etc.). The addressor must also choose the appropriate *setting* (in this case, a professional setting within the given company). Awareness of all this is important because both the type of relationship established between the participants, known as the *tenor*, and the particular setting will significantly affect the shape of the text.

An important aspect of the setting is the physical environment where the communicative event takes place. In writing, awareness of

features in the physical environment is not usually as relevant as it is in spoken communication (e.g., a written job application *versus* an interview), but the time of communication, another element of the setting, may be very important. For instance, time constraints that the writer may have (time/space limitations, deadlines to meet) are also important because they may have an influence on the final product of communication.

As well as deciding on the addressee and the setting, the addressor must make other decisions, such as the most convenient *mode* of communication (written or spoken) and through which *channel* (sound waves, telephones wires, paper, the Internet, etc.). These decisions also influence the shape of the text. In the case of application letters, these decisions are sometimes constrained by a *previous text* that tells the addressor how to apply (e.g., an advertisement of the position that reads, “send a letter of application to”). This also affects the content and form of the text. For instance, it is rather typical of application letters to begin by referring to the source from which the writer learned about the position (e.g., an advertisement in a given newspaper). In addition, the form and content of the message within the text is also affected by the *co-text*. There seem to be some restrictions as to how information should be presented in the text and in what order to make texts more effective from the perspective of the audience. Therefore, for cross-cultural studies it is very important to take the particular rhetorical context in which language

occurs into account.

The addressor must also decide if he/she is going to write seriously, jokingly, or ironically – that is, the addressor must select a *tone*, or key, of communication, which will also affect textual choices. Other decisions involve the *field*, or topic, of communication (which may include, for instance, talking about previous experience in related jobs). This factor is important, since it will have a direct influence on the semantic choices of the text. And depending on the addressee, the addressor will have to decide which *language code* (British English/Peninsular Spanish) and *situational* and *dialectal variety* (formal/informal; standard/some dialectal code) to use to make himself/herself best understood, and thus, to achieve his/her communicative purpose.

Once these decisions have been made, or given, the addressor will have to decide what to say/write – which content to include (*text content*) and how to express it – in what form and layout/format (*text form*). At this point, we should not forget that both the content and the form of the text will be influenced by the addressor's sex, age, personality, emotional state, particular goals – which in this case may be to create the best possible impression on the addressee – previous knowledge of the relevant world (e.g. his familiarity with the company he is going to apply for), previous experience of communicating for similar or other purposes, and sense of

the addressee's expectations.

One important factor in shaping the addressor's knowledge, experience, and ability to communicate for different purposes is the *language*/*writing culture*(s) into which he/she has been socialized. This socialization may have taken place in various environments (or small cultures) such as family, various levels of formal education, friends, workplace, a given *time* in history, and a given *place* in the world. Although each of these small writing cultures operates according to its own norms, values, common practices, and so on, that are learned, they are also likely to interact in complex ways (Connor, 2005) that will affect a writer's writing behaviour, both the process and the product of writing.

Finally, after all the efforts made by the writer to achieve his/her communicative purpose, the *intended purpose* may not be *perceived* by the reader exactly in the same way as the purpose envisioned by the writer. The reader's interpretation will also be affected by other relevant contextual factors (concentration, interest, emotional state, particular goals, etc.).

Although the communication process is surely more complex than what this model represents, it should be emphasized that these factors may and do have an influence on the form and content of a text and should be taken into account in any characterization of texts. Although a complete characterization of all genres in these terms is still lacking, there is

empirical research that supports this theoretical conception of texts as situated genres. This is the way languages and texts should be described, both intra-culturally and cross-culturally.

The application of this theoretical conception of texts to the description of the English language began to flourish in the 1990s with the pioneering work by Swales (1990). This, in combination with insights from corpus linguistics, has driven linguists to analyse large amounts of data in the search for more accurate and reliable descriptions of genres. For instance, Upton (2002) sought to combine the tools of corpus analysis with the specificity of genre analysis in a way that had not been done before to provide a new perspective on a genre, like the fundraising letter, that was not well understood. There is still a need to approach more genres in this and other ways (Moreno, 2003; Connor & Anthony, 2005) to make language descriptions more adequate and useful for fields of application such as language teaching.

As for the cross-cultural description of languages, one added problem involves the issue of corpora comparability, the issue at stake in this chapter. Although important methodological contributions are being made by linguists to describe languages contrastively while benefiting from the powerful tools provided by corpus linguistics (Rabadán, Labrador & Ramón, 2004), considerable work still needs to be done to assure that the corpora are really comparable. For instance, the corpora

used by Rabadán, Labrador and Ramón (2004) was the COBUILD-Bank of English for the English data and the CREA (Corpus de Referencia del Español Actual) for the Spanish data because these were the most comparable and representative of each language at the time. However, these corpora do not meet the methodological requirements for the type of cross-cultural study proposed in this chapter.

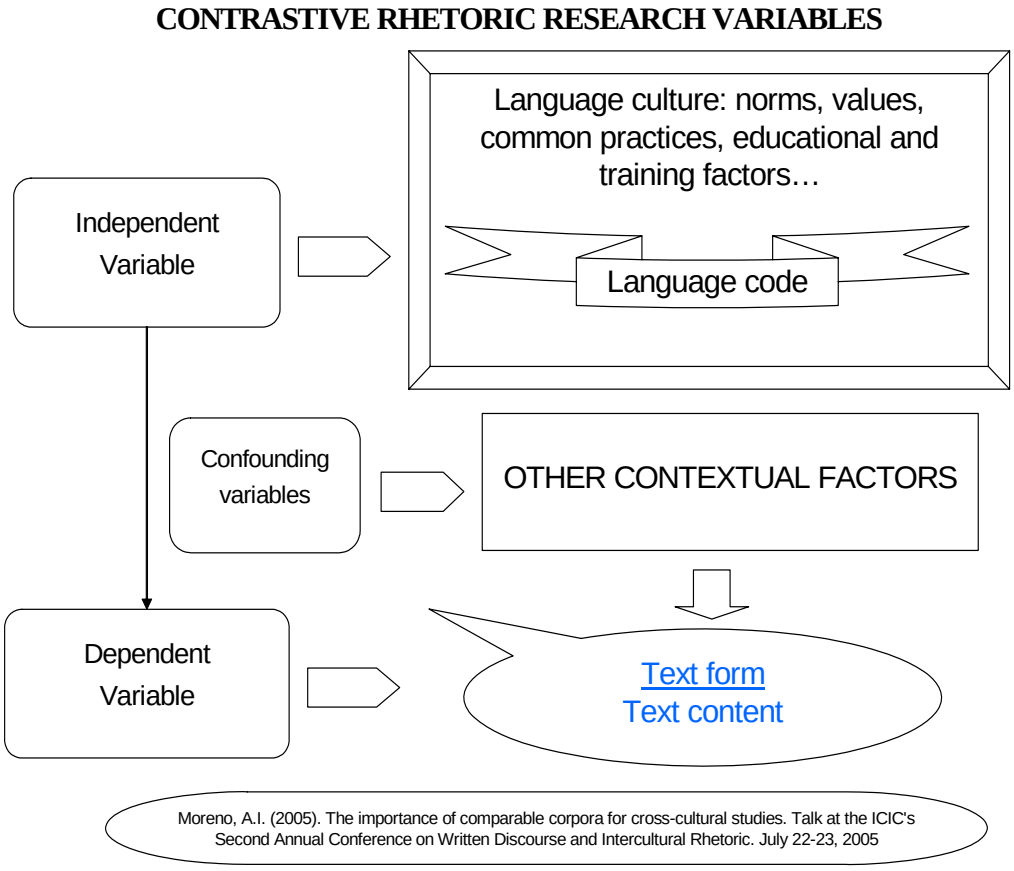
Given the present state of knowledge about language internal variability, arguments for establishing the comparability of the selected subcorpora will need to be more delicate than simply declaring that the selected subcorpora comprise written texts: newspapers, magazines, books and ephemera. One important reason is that newspapers, for instance, include many different genres of texts (news articles, editorials, comment articles, and interviews) which follow different rhetorical conventions. Another reason is that these genres may be represented in such different proportions that it may be impossible to consider the two corpora as statistically comparable. Thus, cross-cultural studies of the type proposed here are very likely to need to develop their own tailor-made corpora, as has been done in studies like Moreno (1996, 1997, 1998, 2004) and the chapter by Suárez and Moreno in this volume.

Research Variables in Cross-Cultural Studies

Since quantitative cross-cultural studies usually compare equivalent rhetorical features of written texts across languages and cultures, they need to be based on comparable written corpora, except for two contextual factors. These two factors are likely to be the *language code* factor, associated with a *writing culture* resulting from the interaction of various small cultures, and the *form and content of the text* factors – i.e., the rhetorical, semantic, and linguistic configuration of texts. Research examines the effect of one (the writing culture expressed through a language code) on the other (form and content of the text). These contrastive research variables are shown in Figure 2.

It may be said that the *language code*, inescapably associated with a writing culture, is the *independent variable* – the one that is manipulated to see how that change affects the shape of language – while *form and content of the text* is the *dependent variable*. It includes the rhetorical/semantic/linguistic features that will be observed and measured to see how they have changed as a function of the writing culture.

Figure 2. Contrastive rhetoric research variables



In order to reach reliable conclusions, the researcher should attempt to maintain constant all other relevant factors affecting the production process – what is known as the *confounding factors*, or variables. The problem is that in descriptive studies, like cross-cultural studies, it is not possible to manipulate the variables. When texts are collected, they are already products. All the possible variables affecting the production process already have a fixed value (e.g., the text is either written or spoken). Therefore, the only way we can perform descriptive

work is to design our corpora very carefully in such a way that all relevant confounding variables are taken into account by statistical control of the sample. If the design is able to maintain constant the values of these confounding factors and manages to include the same proportion of texts representing those values in each sample, we can then say that the two corpora are equivalent to the maximum degree of similarity and each of those factors can be considered as the criteria of comparison that allow us to make a valid comparison.

Table 2 includes a comparable corpus design that met this requirement. This corpus was used by Moreno (1997, 1998) in an English-Spanish contrast of the explicit signalling of various types of causal intersentential relations in research articles on Business and Economics. For instance, the independent variable was the language code (with two possible values: English versus Spanish), and the dependent variable in Moreno (1998) was the explicit signalling of premise-conclusion intersentential relations, which was broken down into more specific dependent variables.

Table 2. Similarity constraints established for the design of Moreno's (1998) English-Spanish comparable corpora of research articles on business and economics (adapted from Moreno, 1996, p. 162)

Criteria of comparison	**Value of prototypical feature perceived as a constant across the two corpora**	**N of texts in each independent corpus**
Text form	Scientific exposition	36
Genre	Research article	36
Mode	Written language	36

Participants Writers Targeted readers	 Researchers, professors and professionals in business and economics Researchers, professors, advanced students, top executives, politicians	 36
Situational variety Dialectal variety	Formal Standard	36
Tone	Serious	36
Channel	Graphical substance	36
Format features Length Intertextuality Visual features	 2.000-16.000 words of core text Including reference to other texts Including graphs, tables, drawings, footnotes, appendixes, typographical distinctions to indicate sections	 36
Point of view	Objective	36
Global communicative event	Sharing results from research	36
Setting	An office, a library,...	36
General purpose of communication	Writer's viewpoint: To persuade the readers to share the writer's viewpoint Reader's viewpoint: To improve one's knowledge about a given field of research	36
Global rhetorical strategy	Demonstrating a theory Discussing the advantages of: • applying a given model • a given business practice Analysing the reasons for a given situation Proving the accuracy of a prediction Evaluating the solution given to a situation	36
Overall subject-matter or topic Academic discipline	Business and Economics Marketing-management Economics-finance	36 18 18
Level of expertise	Expert writers	36
Textual unit of analysis	Complete texts	36
Global superstructure	Introduction-Procedure-Discussion More variable superstructures: Problem-Analysis-Solution Situation-Explanation Situation-Analysis-Forecast Problem-solution-Evaluation	11 25
Predominant text types (Depending on the focus of each section in the superstructure of the article)	Argumentation Exposition Description	36

The first column in Table 2 shows the criteria of comparison that helped the researcher to make similarity judgments between the two corpora. As can be seen, these criteria of comparison correspond to the various relevant contextual factors, or confounding variables, that were hypothesized to have an influence on the form and content of the message. The second column shows the particular value each factor took on that was kept constant for both corpora. The corpora can be considered comparable, or equivalent, in all those respects.

It would be possible to suggest that the criterion referred to as *genre* could well serve as a comprehensive criterion of comparison since it usually restricts the values of the other contextual factors. For instance, a research article (a possible value of the *genre factor*) is normally expressed in the written mode (a possible value of the *mode* factor) through graphical substance on paper (a possible value of the *channel* factor). It normally takes the form of scientific exposition; the participants are usually restricted to researchers, professors and professionals, or advanced students; these texts are usually written in a formal and standard variety; their tone is usually serious; they have a recognizable format in terms of length, intertextuality and visual features; they tend to be presented as objective in viewpoint; the global communicative event in which they are framed is that of sharing results from research. The general purpose of communication, from the point of view of the writer, is usually

to persuade readers to share the writer's viewpoint, and from the readers' perspective, to improve knowledge about a given field of research. A research article may also follow recurrent global rhetorical strategies (demonstrating a theory, discussing the advantages of applying a given model or a given business practice, analysing the reasons for a given situation, proving the accuracy of a prediction, evaluating the solution given to a problem). The predominant text types, depending on the precise section of the article, seem to be argumentation, exposition, and description. They are usually written by expert writers.

Even if the above values are constant, there are still a number of factors which may escape this control by the *genre* factor. For example, the genre we recognize as the academic research article is actually generated within differing *disciplinary cultures*. These disciplinary cultures may have their own conventions of writing this kind of text. Since the disciplinary culture is associated with what is known as the *topic* or *field* factor, Moreno's study took this variable into account in the sense that it made both corpora balanced in terms of *topics*. Accordingly, the resulting sample in each corpus consisted of 18 research articles about marketing-management and 18 research articles about finance-economics.

Although Moreno's cross-cultural study examined a text-rhetorical feature in the entire research article, the study conjectured that this text-rhetorical feature might also vary as a function of the particular section

within the research article. The study attempted to control for the *superstructure* factor. Thus, 11 research articles in each corpus followed the overall pattern of Introduction-Procedure-Discussion, and 25 research articles in each corpus showed more variable superstructures: Problem-Analysis-Solution; Situation-Explanation; Situation-Analysis-Forecast and Problem-Solution-Evaluation.

Another important factor that would escape the control of the *genre* criterion is the temporal factor (see *temporal coordinate* in Figure 2.1). This diachronic factor should also be taken into account because genres are dynamic entities. Any cross-cultural study should also specify the temporal coordinates. Since Moreno's study attempted to do a cross-cultural characterization of the most recent research articles at the time of the research, the sample texts were restricted to research articles written between 1990 and 1993.

It is impossible to make constant certain factors (e.g., the addressor). This is a complex factor comprising other factors (age, sex, experience in writing, maturity, personality, etc.) which may affect the form of the message even within the same writing culture. Every writer has idiosyncrasies, and it is probably impossible to make two corpora similar in this respect. Still, if we conduct cross-cultural studies in an attempt to capture general tendencies of particular writing cultures, it seems that the best solution is to draw on corpus linguistics. We can

design cross-cultural corpora consisting of large numbers of texts written by a great variety of authors, selected randomly to represent the targeted population of texts. This way, the possible effect of idiosyncrasies is diluted within the multitude. If, based on previous research, the researcher has some ground to think that a given factor (e.g., gender) is highly influential in some respect of the rhetorical configuration of the texts, it should be taken into account in such a way that the two corpora contain the same proportions of texts written by the two genders. We need to control that factor statistically by means of stratified sampling.

Finally, although developing comparable corpora in this controlled way already guarantees that many of the relevant contextual factors are taken into account, using traditional tools of corpus linguistics to assist analysis (e.g., concordances) still carries the danger of ignoring the precise rhetorical context in which language features occur. Therefore, more complex, analytical, computer-based tools should be developed (e.g., tagging the corpora for rhetorical moves and functions) before we can reasonably exploit the power of corpus analysis tools (Connor and Anthony 2005). Until that happens, cross-cultural linguists will have to continue their traditional manual analysis of texts in the search for rhetorical patterns of each language.

Conclusion

Comparable corpora in cross-cultural research on written discourse are important. If we do not design our corpora carefully, we may not account for confounding factors and eventually may not be able to say anything reliable about the possible effect of the independent variable (the writing culture factor associated to a language code) on the dependent variable(s). If the confounding variables are left uncontrolled and we observe cross-cultural differences in relation to a given rhetorical feature, we will not be able to attribute them to the effect of the writing culture, or language code, because they may have been due to the effect of some confounding variable.

If further CR studies of question type 1 (from Table 2.1) between any pair of languages make explicit the criteria of comparison at the level of corpus design, we may eventually build a dynamic model of Rhetorics in contrast. Such CR studies would provide us with stable foundations upon which to build further applied studies, such as language learning/teaching and translation.

It is important to highlight the fact that once differences (and similarities) in discourse structure and rhetorical features (question type 1: Whether the imputed cross-cultural differences in the rhetorical configuration of texts actually exist) are identified, further qualitative

research (question type 2: If they exist, which cultural or educational factors may help to account for such differences [e.g., values, norms, learning processes and educational trends]) should attempt to explain the sources of that variation by referring to cultural features of the two writing cultures. This is another important way in which cross-cultural studies can attempt to take context into account. Having discovered cross-cultural differences in rhetorical patterns of texts, the next logical step is to pinpoint which specific aspects (e.g., values, norms, common practices, and learning processes) of the writing cultures are responsible for a given variation in rhetorical behaviour. That is, not only awareness of the differences (and similarities) but also the reasons for such divergence would be helpful in applied fields such as the teaching of writing in English as L2. However, as Connor (2004) highlights, “teachers must keep in mind that no one needs to be held hostage by language and culture; students can be taught to negotiate conflicting rhetorical structures to their advantage” (Connor, 2004, p. 271).

It is also important to emphasize that the relevance of researching the first two types of questions should be established by reference to studies or teaching/learning experiences where difficulties with discourse structure are identified. Otherwise, what would be the point? That is why rigorous studies that answer questions of type 3 (Which precise difficulties with discourse structure and other rhetorical features do second language

learners from a given non-English writing culture experience when writing in English as an L2?) are also important. Once linguists, discourse analysts, and rhetoricians have provided answers to the first three questions (which difficulties with discourse structure and rhetorical features are experienced; whether the imputed differences exist; and which cultural or educational factors may account for them), further studies may aim to research the fourth type of question. These studies will answer whether difficulties experienced with discourse structure and other rhetorical features by L2 learners of English are attributable to interference (or negative transfer) from the first language.

Finally, the framework of comparison that Moreno (1998) used, and that Connor and Moreno (2005) proposed for cross-cultural study around the concept of genre, seems to be a valid framework as long as the given genre is comparable in the two writing cultures compared. However, there may be cases where this will not hold. For instance, there may be differences in the frequency of use of genres to achieve similar purposes of communication. (As Fusari, 2005, showed, for example, direct mail fundraising letters are not as frequent in Italy as they are in the U.S., nor are the causes for which money is raised the same). If corpora are not selected carefully, it will be more difficult to determine which contextual factor is responsible for the possible differences in rhetoric. There may also be cases of genres that did not exist in one language (e.g., Spanish

conditions of sales) before translations of English ones appeared. We may find that this genre written in Spanish follows the English genre expectations. Such cases will not lend themselves to interesting comparisons from a CR point of view.

Acknowledgements

This chapter is part of a research project financed by the Spanish Ministry of Science and Technology – Plan Nacional de I+D+I (2000-2003) Ref: BFF2001-0112, entitled *Contrastive Analysis and Specialized English-Spanish Translation: Applications and Tools (ACTRES).* I am also grateful to Dr. Ulla Connor, Dr. William Rozycki, and an anonymous scholar for their comments on an earlier version of this chapter.

References

Chesterman, A. (1998). *Contrastive functional analysis*. Amsterdam: John Benjamins.

Connor, U. (2002). New directions in contrastive rhetoric. *TESOL Quarterly, 36*, 493-510.

Connor, U. (2004). Intercultural rhetoric research: Beyond texts. *Journal of English for Academic Purposes, 3*, 291-304.

Connor, U. (2005, July). *Intercultural rhetoric – Further defined.* Paper presented at the Indiana Center for Intercultural Communication's Second Annual Conference on Intercultural Rhetoric and Written Discourse Analysis, Indianapolis, IN.

Connor, U., & Anthony, M. A. (2005, July). *Semantic variation in rhetorical appeals of fundraising letters.* Paper presented at the "Research on Fundraising Letters: Focus on Research Methods" symposium at the 14th World Congress of Applied Linguistics, Madison, WI.

Connor, U., & Moreno, A. I. (2005). Tertium comparationis: A vital component in contrastive rhetoric research. In P. Bruthiaux, D. Atkinson, W. Eggington, W. Grabe & V. Ramanathan (Eds.), *Directions in applied linguistics: Essays in honor of Robert B. Kaplan* (pp. 153-164). Clevedon, UK: Multilingual Matters.

Cook, G. (1989). *Discourse*. Oxford: Oxford University Press.

Fusari, S. (2005, July). *A corpus study of philanthropic direct mail.* Paper presented at the Indiana Center for Intercultural Communication's Second Annual Conference on Intercultural Rhetoric and Written Discourse Analysis, Indianapolis, IN.

Halliday, M. A. K., & Hasan, R. (1989). *Language, context, and text:*

Aspects of language in a social-semiotic perspective. Oxford: Oxford University Press.

Hymes, D. H. (1968). The ethnography of speaking. In J. A. Fishman (Ed.), *Readings in the sociology of language* (pp. 99-138). The Hague, Paris: Mouton.

Jakobson, R. (1960). Closing statement: Linguistics and poetics. In T. A. Sebeok (Ed.), *Style in language* (pp. 350-377). Cambridge, MA: Technology Press of Massachusetts Institute of Technology.

Kaplan, R. B. (1987). Cultural thought patterns revisited. In U. Connor & R. B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text* (pp. 9-21). Reading, MA: Addison-Wesley.

Kaplan, R. B. (1966). Cultural thought patterns in inter-cultural education. *Language Learning: A Journal of Applied Linguistics, 16*, 1-20.

Maslow, A. H. (1954). *Motivation and personality*. New York: Harper.

Moreno, A. I. (1996). *Estudio Contrastivo Inglés-Español de la Expresión de las Relaciones de Coherencia Causal Interoracional: el Artículo Académico sobre Economía y Empresa. Tesis doctorales 1995*. Published doctoral dissertation, Universidad de León, León, Spain.

Moreno, A. I. (1997). Genre constraints across languages: Causal metatext in Spanish and English RAs. *English for Specific Purposes, 16*, 161-179.

Moreno, A. I. (1998). The explicit signaling of premise-conclusion sequences in research articles: A contrastive framework. *Text, 18*,

545-585.

Moreno, A. I. (2003). Matching theoretical descriptions of discourse and practical applications to teaching: The case of causal metatext. *English for Specific Purposes, 22*, 265-295.

Moreno, A. I. (2004). Retrospective labelling in premise-conclusion metatext: An English-Spanish contrastive study of research articles on business and economics. *Journal of English for Academic Purposes, 3*, 321-339.

Rabadán, R., Labrador, B., & Ramón, N. (2004). English-Spanish corpus-based contrastive analysis: Transnational applications from the ACTRES Project: PALC 2003. In B. Lewandowska-Tomaszczyk (Ed.), *Practical applications in language and computers* (pp. 141-152). Frankfurt: Peter Lang.

Reid, J. M. (1988). *Quantitative differences in English prose written by Arabic, Chinese, Spanish, and English students*. Unpublished doctoral dissertation, Colorado State University, Fort Collins, CO.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge, UK: Cambridge University Press.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Cambridge, MA: Harvard University Press.

Upton, T. A. (2002). Understanding direct mail letters as a genre. *International Journal of Corpus Linguistics, 7*, 65-85.